

# Hybrid coding of video with spatio-temporal scalability using subband decomposition

Marek Domanski\* , Adam Luczak, Slawomir Mackowiak, Roger Swierczynski

Poznań University of Technology, Institute of Electronics and Telecommunications,  
Poznan, Poland

## ABSTRACT

The paper deals with scalable coding of video with SDTV or HDTV resolution. A new technique of scalable coding is proposed for bitrates of about 3–10 Mbps. The technique has been implemented for BT.601 resolution and progressive scan, therefore problems related to an interlaced scan are omitted here. The goal is to improve spatial scalability of MPEG-2 by introducing spatio-temporal scalability. The technique proposed needs less coding overhead than in MPEG-2 spatially scalable scheme and an enhancement layer bitstream with its bitrate not less than the bitrate in a base layer. The solution proposed in the paper is based on both temporal and spatial resolution reduction performed for data transmitted in a base layer. The temporal resolution reduction is obtained by placing each second frame (B-frame) in the enhancement layer. The enhancement layer includes also high-frequency spatial subbands from other frames. A variant of the system based on three-dimensional spatio-temporal analysis is also described. In both cases the assumption is that a base layer is fully MPEG-2 compatible.

**Keywords:** Spatial scalability, spatio-temporal scalability, MPEG-2, subband decomposition

## 1. INTRODUCTION

Scalable or hierarchical coders produce two bitstreams:

- base layer bitstream which represents low resolution/quality pictures,
- enhancement layer bitstream which provides additional data needed for reproduction of pictures with full resolution/quality.

An important feature is that a base layer bitstream can be decoded independently from an enhancement layer. Therefore low-level terminals are able to decode only the base layer bitstream in order to display low-level pictures. The functionality of scalability is also important for error-resilient video transmission where base layer packets are well protected against transmission errors or packet losses while the protection of the enhancement layer is lower. A receiver is able to reproduce at least low-level pictures if quality of service decreases.

MPEG-2 video compression standard<sup>1,2</sup> established four types of scalability: spatial, temporal, SNR and data partitioning. Among them, spatial scalability is of particular interest because of its prospective broad applications. Unfortunately, spatially scalable systems proposed by MPEG-2 coding standard is inefficient. Recently proposed MPEG-4 has proposed no substantial improvement. Therefore there were many attempts to improve the scheme of spatial scalability by application of subband decomposition<sup>3-7</sup>. The idea is to split each image into four spatial subbands. The subband of lowest frequencies constitutes a base layer while the other three subbands are jointly transmitted in an enhancement layer. Nevertheless this approach often leads to allocation of much higher bitrates to a base layer than to an enhancement layer which is disadvantageous for practical applications.

In order to avoid the above mentioned problem, spatio-temporal scalability is proposed<sup>13</sup>. Here, a base layer corresponds to pictures with reduced both spatial and temporal resolution. An enhancement layer is used to transmit the information

---

\* Correspondence: Piotrowo 3A, 60-965, Poznań, Poland, Telephone: +48 61 8782762; Fax: +48 61 8782572  
E-mail: domanski@et.put.poznan.pl

needed for restoration of the full spatial and temporal resolution. The assumption is that the base layer is fully MPEG-2 compatible.

The paper deals with scalable coding of video with SDTV or HDTV resolution. A new technique of scalable coding is proposed for bitrates of about 3 – 10 Mbps. The technique has been implemented for BT.601<sup>8</sup> resolution and progressive scan, therefore problems related to an interlaced scan are omitted here.

## 2. MPEG-2 SPATIAL SCALABILITY

MPEG-2 spatial scalability<sup>1,2</sup> is based on pyramid decomposition (Fig. 1). Unfortunately, it is inefficient because the number of pixels to be encoded is increased as compared to the nonscalable scheme.

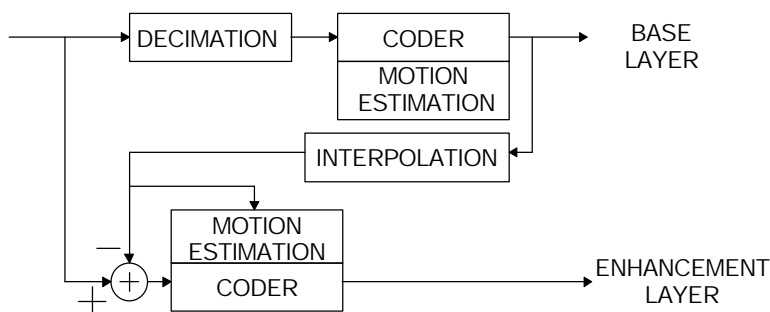


Figure 1: Spatially scalable coder as proposed by MPEG-2 standard.

The experiments made with test video sequences show that MPEG-2 spatially scalable scheme exhibits substantial disadvantages:

- Bitstream increased by 60% - 75% as compared to MPEG-2 single layer coding.
- Similar performance as simulcast coding.

Therefore such scalable systems are hardly used in most applications.

The conclusion is that there exists a need to improve performance of scalable video codecs.

## 3. SPATIALLY SCALABLE CODERS WITH SUBBAND DECOMPOSITION

A very straightforward proposal is to use subband decomposition in order to obtain spatial scalability. As mentioned above, a number of such solutions has been already proposed. Their arrangement is as follows. The video sequence is decomposed frame by frame by a FIR filterbank. This analysis results in decomposition of the video sequence into four spatial subbands (Fig. 2) which are assigned to the base layer and the enhancement layers.

Standard solution is as follows:

*Base layer:* Subband LL,  
MPEG-2 encoded

*Enhancement layer:* Subbands LH, HL, HH,  
Full-frame motion compensation,  
DCT-based in-band hybrid coding.

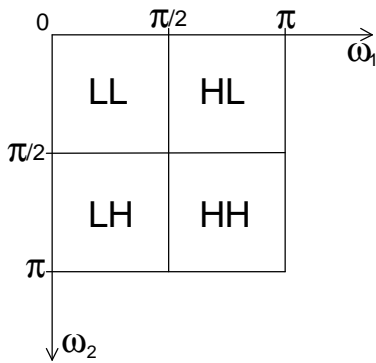


Figure 2: Subband decomposition into four spatial subbands. Separable filter banks are used.

Proper motion compensated scalable coding is a problem that is closely related to subband coding of video where it has been already found that in-band motion estimation and compensation usually leads to poor coding efficiency.

Despite of the type of motion compensation, the above system exhibits a substantial disadvantage. It is caused by the fact that bitstream does not decrease linearly with the number of pixels in a frame. The experiments show that the base layer bitstream is usually much larger than the enhancement layer bitstream<sup>10</sup>. Such a system is deprived of practical usefulness if the enhancement layer bitstream is much less than the base layer bitstream. For example, a system where only 20-30% of data are less protected against packet losses brings not too much improvement as compared to a single layer system.

The solution seems to be application of a combination of spatial scalability with another type of scalability.

#### 4. SPATIO-TEMPORAL SCALABILITY

Our idea is to combine spatial and temporal scalability.

Following assumptions have been made for the solutions proposed:

- The base layer coder is fully MPEG-2 - compatible.
- The coder produces similar bitstreams in the base layer and the enhancement layer.

Base layer represents video with reduced both temporal and spatial resolutions. The Enhancement layer is used to transmit the information needed for restoration of the full spatial and temporal resolution. Two coder versions have been considered by the authors.

Systems of the first type incorporate B-frames into enhancement layer. Base layer consists of the subband LL from each even frame. Enhancement layer includes subbands LH, HL, HH from each even frame as well as each whole odd frame which is assumed to be a B-frame (Fig. 3).

<i>Base layer:</i>	I		B		P		B		P		B		P
<i>Enhancement layer:</i>	I	B	B	B	P	B	B	B	P	B	B	B	P

Figure 3: Picture types in both layers of the first system.

Another proposal is to use systems with three-dimensional subband analysis. The input video sequence is analyzed in a three-dimensional (3-D) separable filter bank, i.e. there are three consecutive steps of analysis: temporal, horizontal and vertical. Temporal analysis results in two subbands  $L_t$  and  $H_t$  which are partitioned into four spatial subbands (LL, LH, HL and HH) each. For spatial analysis, both horizontal and vertical, separable filters are used. The three-dimensional analysis results in eight spatio-temporal subbands. Three high-spatial-frequency subbands (LH, HL and HH) in the high-temporal-frequency subband  $H_t$  are discarded as they correspond to the information being less relevant for the human visual system. Therefore five subbands are encoded:

- In a base layer - the spatial subband LL of the temporal subband  $L_t$ .
- The enhancement layer includes the spatial subbands LH, HL and HH from the temporal subband  $L_t$  and the spatial subband LL of the temporal subband  $H_t$ .

The base layer is produced by an MPEG-2 motion compensated coder.

Both variants will be discussed further.

## 5. CODING EXPLOITING INTER-SUBBAND CORRELATION

In both above described systems, a video sequence at input to the spatial analysis filter bank is a sequence of pictures similar to those in the source video sequence. Usually pictures showing natural scenes exhibit very high cross-correlation between sub-images corresponding to different subbands of the spatial spectrum. Therefore a technique originally successfully applied for still images<sup>12</sup> can be adopted for efficient encoding of the subbands HL, LH and HH (Fig. 4). This technique exploits an observation that images of natural scenes exhibit power spectra which decay more or less uniformly for increasing frequencies. Therefore, the portions of the high-frequency sub-images where signal values are substantial correspond usually to the portions of the low-frequency sub-image where some changes occur. Most of the portions of the high-frequency sub-images exhibit only small signal values. Therefore a quantizer with a dead-zone produces a lot of zeros. We need not to transmit all of them and we can use the low-frequency sub-image to identify the positions of the "active" portions of the high-frequency sub-images.

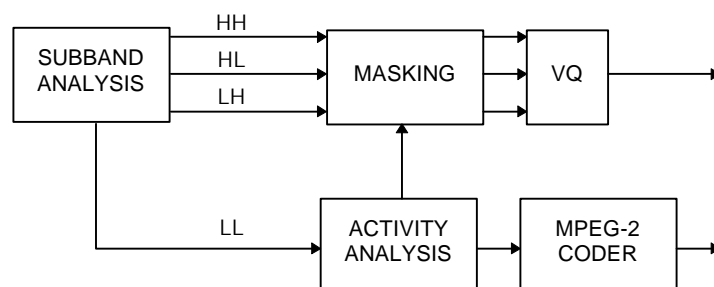


Figure 4: Coding exploiting inter-subband correlation.

## 6. INCORPORATION OF B-FRAMES INTO ENHANCEMENT LAYER

It is assumed that video sequences are encoded using the group of pictures (GOP) structure that consists of one or three B-pictures between two consecutive I- or P-pictures. Discarding B-frames and decoding only I- and P-frames is temporal decimation. Some kind of poor low-pass filtering is made due to quantization performed in the coder. In our proposal we enclose each odd-numbered frame is included into enhancement layer. These frames are obviously B-frames due to the above assumptions. The decoder structure is shown in Fig. 5.

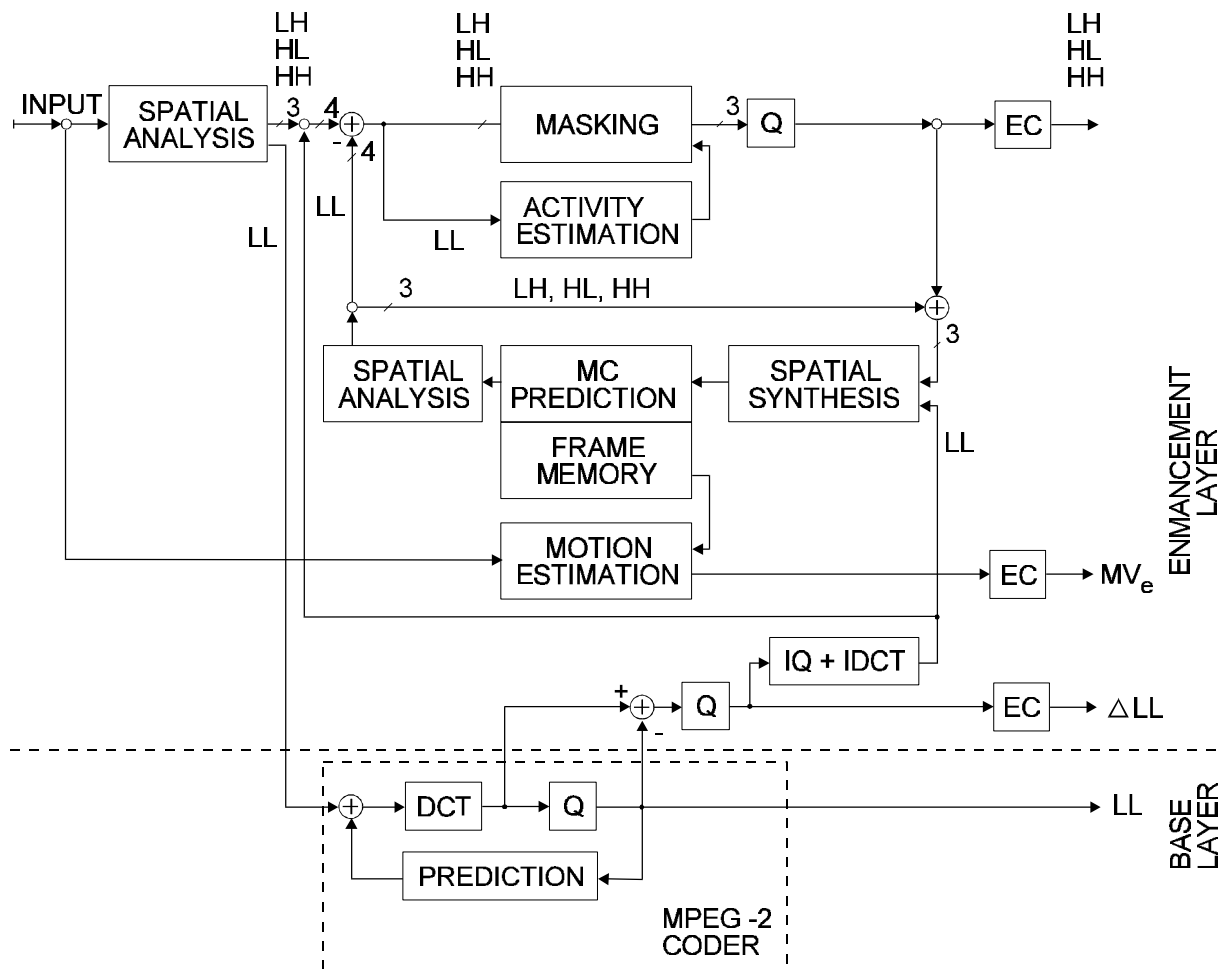


Figure 5: Coder that exploits temporal decomposition according picture type.

The base layer bitstream is produced in a standard MPEG-2 coder that processes each second frame with reduced spatial resolution. The experiments have been made with 50Hz  $720 \times 576$  progressive sequences which correspond to 25Hz SIF progressive sequence in the base layer. The bitstream produced in the base layer coder is an MPEG-2 bitstream that includes also motion vectors estimated with half-pel accuracy for SIF images. In the case of HDTV processing the base layer would encode 576-line images possibly of 16:9 picture aspect.

The enhancement layer coder is not MPEG-2 compatible but exploits several functions implemented in MPEG-2 coders. The main feature is full-frame motion estimation and compensation. In order to be able to do it, spatial synthesis must be done, i.e. four subbands (including the LL subband that constitutes the base layer) are synthesized into one signal. Then motion estimation and compensation is done. In order to speed up motion estimation the base layer motion vectors are used as initial guess. Note that their accuracy is full-pel in a full-size image. After motion compensation the pictures are split into four spatial subbands again in order to encode them individually.

The enhancement layer encoder produces four bitstreams:

1. HL subband,
2. LH subband,
3. HH subband,
4.  $\Delta LL$  – corrections to the LL subband,
5.  $MV_e$  – Full-frame motion vectors.

1 – 3. Many different compression techniques are applicable to these subbands. The experiments have been done with the technique described above. This technique exploits mutual dependencies between subbands, or more exactly, between

prediction errors in individual subbands. The data obtained are then encoded using Huffman coders. The tables of codes are roughly estimated.

Another possibility considered was context-based geometric vector quantization.

4.  $\Delta LL$  is the most difficult signal to encode. This signal is needed to improve quality of the LL subband because full frame image quality is very sensitive to LL image quality<sup>14</sup>. This fact is well-known from subband coding of images. Unfortunately,  $\Delta LL$  bitrate grows rapidly as bitstream in the base layer decreases. A possible solution to overcome this problem would be to encode  $\Delta LL$  together with other enhancement layer signals.

5.  $MV_e$  is a bitstream which encodes motion vector refinements. In the decoder, motion vectors are restored by use of the base layer motion vectors and the refinements  $MV_e$ .

The codec has been implemented as software written in C++ language (about 13 000 code lines). Currently the algorithm is being tuned.

The system aimed at 7 Mbps total bitrate has been tested for 30-32 dB PSNR in the luminance component. The scalability overhead is currently about 40 - 50 % of the MPEG-2 single layer bitrate. Further improvement can be expected after algorithm tuning. The base layer bitrate is of the same order as the enhancement layer bitrate.

## 7. SYSTEMS WITH THREE-DIMENSIONAL SUBBAND ANALYSIS

The input video sequence is analyzed in a three-dimensional (3-D) separable filter bank, i.e. there are three consecutive steps of analysis: temporal, horizontal and vertical. For temporal analysis, very simple linear-phase two-tap filters are used similarly as in pure three-dimensional subband coding<sup>9-11</sup>

$$H(z) = 0.5(1 \pm z^{-1}),$$

where "+" and "-" correspond to low- and high-pass filters, respectively. This filter bank has a very simple implementation and exhibits small group delay (half of sampling period) resulting in small system response times which are very critical for many applications in interactive multimedia.

Temporal analysis results in two subbands  $L_t$  and  $H_t$  which are partitioned into four spatial subbands (LL, LH, HL and HH) each. For spatial analysis, both horizontal and vertical, separable nonrecursive half-band linear-phase filters in polyphase implementation are used. Linear phase is assumed here because motion vectors estimated in the LL subband of the  $L_t$  temporal subband are used for motion compensation in the LL subband of the  $H_t$  temporal subband.

The three-dimensional analysis results in eight spatio-temporal subbands. Three high-spatial-frequency subbands (LH, HL and HH) in the high-temporal-frequency subband  $H_t$  are discarded as they correspond to the information being less relevant for the human visual system. Although it reduces PSNR, e.g. in the intraframe mode to about 32-33 dB, it has small influence on subjective quality of the decoded video.

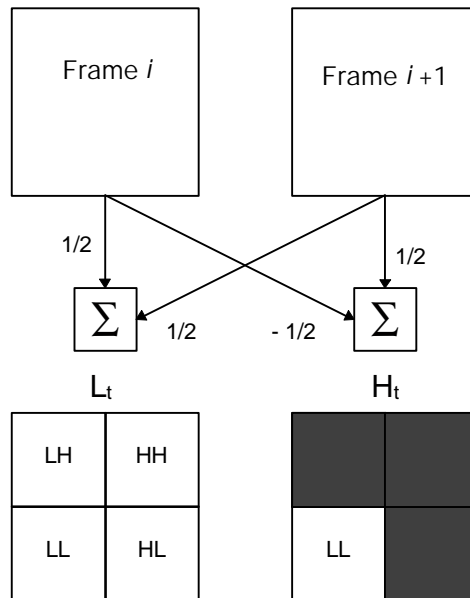


Figure 6: Three-dimensional subband analysis.

Therefore five subbands are encoded (see Fig. 6):

- In a base layer - the spatial subband LL of the temporal subband  $L_t$ .
- The enhancement layer includes the spatial subbands LH, HL and HH from the temporal subband  $L_t$  and the spatial subband LL of the temporal subband  $H_t$ .

The base layer is produced by an MPEG-2 motion compensated coder.

The coder structure has been chosen after some experiments with various variants of coding algorithm. The experiments have been made for progressive format of video, and interlaced video is not considered in this paper.

Two subbands are encoded using MPEG-compatible DCT-based coders. The base layer is produced by an MPEG-2-compatible coder which processes the subband LL from the temporal subband  $L_t$ . Motion estimation, the most computationally demanding task is performed in this subband. The motion vectors obtained are used also for the LL subband from the temporal subband  $H_t$  which is encoded using an MPEG-2-compatible motion-compensated coder. Nevertheless this coder does not include a motion estimator.

Motion compensation in the LL subband from the temporal subband  $H_t$  is not very efficient. The experiments made resulted in a conclusion that the bitrate reduction caused by motion compensation usually does not exceed 20%.

As mentioned above, the LL/ $L_t$  subband is encoded using an MPEG-2 coder. In order to obtain high compression ratio for the base layer, rough quantization of the DCT coefficients is employed, e.g. with the quantization coefficient of about 8-16. Such quantization would decrease the quality of the full resolution image too much. Therefore some corrections for non-zero valued DCT coefficients are additionally transmitted in the enhancement layer. These corrections are encoded using Huffman codes and their locations are defined by the locations of the non-zero coefficients from the base layer.

The experiments prove that motion compensation in the three high-spatial-frequency subbands (LH, HL and HH) in the low-temporal-frequency subband  $L_t$  is inefficient. This corollary is similar to that from [6]. Therefore these subbands are encoded without motion compensation. In order to exploit mutual dependencies between subbands, an original technique previously proposed for still images<sup>12</sup> is used. The pixels which do not correspond to the "active" portions of the LL/ $L_t$  subband are discarded. The remaining pixels are quantized by a nonlinear quantizer and encoded using DPCM and variable-length coding.

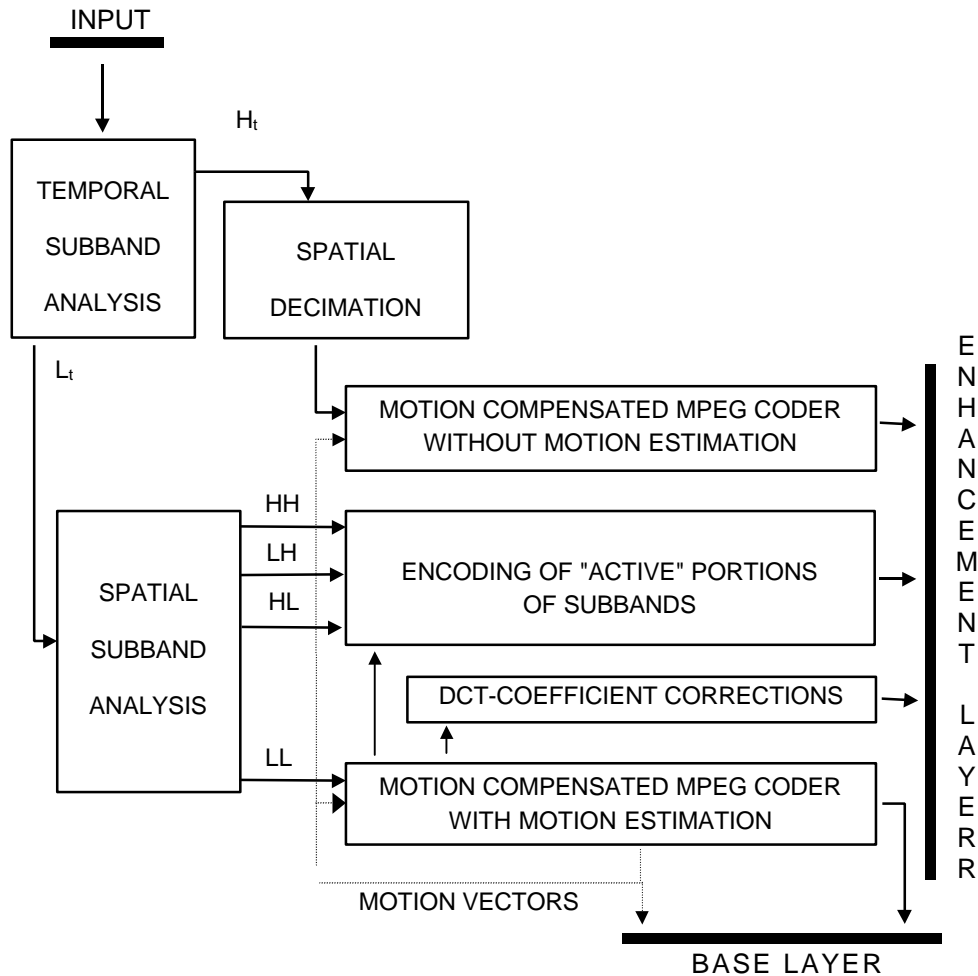


Figure 7: Coder structure.

Preliminary estimations show that the performance of this system is similar to the previous one.

## 8. CONCLUSIONS

The paper describes a new scalable video coder. A new feature is application of three-dimensional subband decomposition in a scalable MPEG-compatible coder.

The structure of the coder has been chosen according to the results of preliminary experiments with test video sequences. Enormous work has been already done by preparation the software that implements the video codec. This software is currently being tested but some promising preliminary experimental results have been already obtained. Further work will include experiments which should estimate the properties of the coder. Moreover the coder needs optimization and tuning by adjustment of its parameters.

## ACKNOWLEDGEMENT

The authors express their sincere thanks to Prof. H.G. Musmann and Mr. U. Benzler from the University of Hannover, Germany for their help, advice and fruitful discussions.

## REFERENCES

1. ISO/IEC International Standard 13818, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information*.
2. B.G. Haskell, A. Puri, A. N. Netravali, *Digital Video: an Introduction to MPEG-2*, Chapman & Hall, 1997.
3. T. Tsunashima, J. Stampleman, V. Bove., "A scalable motion-compensated subband image coder", *IEEE Trans. on Communication*, **42**, pp. 1894-1901, 1994.
4. U. Benzler, "Scalable Multiresolution Video Coding Using Sub-Band Decomposition", *First Int. Workshop on Wireless Image/Video Communication*, Loughborough, pp. 109-114, 1996.
5. H.Gharavi, W.Y.Ng, "H.263 Compatible Video Coding and Transmission", *First Int. Workshop on Wireless Image/Video Communication*, Loughborough, pp. 115-120, 1996.
6. B. Uz, M. Vetterli, D. J. LeGall, "Interpolative multiresolution coding of advanced television with compatible subchannels", *IEEE Transactions on Circuits and Systems for Video Technology*, **1**, 1991.
7. H. Gharavi, "Subband coding of video signals", chapter 6 in: *Subband Image Coding*, edited by J. W. Woods, Kluwer, Boston, 1991.
8. Recommendation ITU-R BT 601-4, *Encoding parameters of digital television for studios*.
9. R.Ohm, "Three-dimensional subband coding with motion compensation", *IEEE Trans. Image Proc.*, **3**, pp.559-571, 1994.
10. M. Domanski, R. Swierczynski, "3-D subband coding of video using recursive filter banks", *Signal Processing VIII: Theories and Applications*, Trieste, pp. 1359-1362, 1996.
11. Ch. Podlichuk, N. Jayant, N. Farvardin. "Three-dimensional subband coding of video", *IEEE Trans. Image Proc.*, **4**, pp.125-284, 1995.
12. M. Domanski, R. Swierczynski, "Subband coding of images using hierarchical quantization", *Signal Processing VII: Theories and Applications*, Edinbrough, pp. 1218-1221, 1994.
13. M. Domanski, A. Luczak, S. Mackowiak, R. Swierczynski, "Hybrid coding of video with spatio-temporal scalability using subband decomposition", *Signal Processing IX: Theories and Applications*, Rhodes, pp. 53-56, 1998.
14. J. Woods (ed.), *Subband image coding*, Kluwer, Boston , 1991.